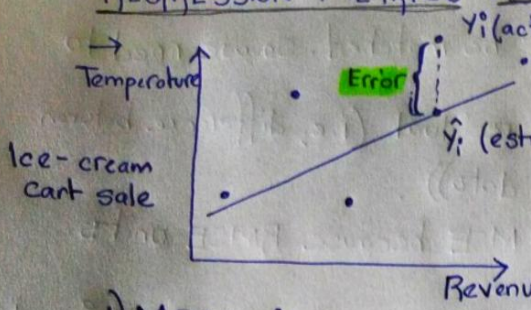
REGRESSION METRICS : HOW TO ASSESS MODEL PERFORMANCE

DAY 7



Residual (Error) = $\hat{y}_i - y_i$
= prediction - Actual

if the model is perfect, residual = 0.

In this case, we have to find residual for every point.

1) MEAN ABSOLUTE ERROR (MAE)

→ MAE is obtained by calculating the absolute differences between the model predictions and the true (actual) values.

→ MAE is a measure of the average magnitude of error generated by regression model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

→ If we do not take absolute, sum will become zero.

→ MAE is calculated by following steps:

1. Calculate the residual of every data point.
2. Calculate the absolute value (to get rid of sign).
3. Calculate the average of all residuals.

→ Even after absolute value if MAE is zero, this indicates that model prediction are perfect.

2) MEAN SQUARE ERROR (MSE)

→ MSE is very similar to MAE but instead of using absolute values, square of the difference between the model predictions and the training dataset (true values) is being calculated.

→ MSE values are generally larger compared to MAE since the residual are being squared.

→ In case of data outliers, MSE will become much larger to MAE.

→ In MSE, error increase in quadratic fashion while error increase in proportional fashion in MAE

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

→ MSE is calculated in following steps:

1) Calculate the residual for every data point.
2) Calculate the square value of residual
3) Calculate the average of results from step #2.

→ In MSE, since the error is being squared, any predicting error is being heavily penalized.

3) Root Mean Square ERROR (RMSE)

→ In MSE, we square the error so unit cannot be matched. So we need to square root (cancelling the effect).

→ RMSE represents the standard deviation of residual (i.e, difference between the model predictions and true value (training data)).

→ RMSE can be easily interpreted compare to MSE because RMSE units match units of the output.

→ RMSE provide an estimates of how large the residual are being dispersed

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

→ RMSE is calculated by following steps –

1. Calculate the residual of every data point
2. Calculate the square value of residual.
3. Calculate the average of square residual.
4. Obtain the square root of residual.

→ Standard deviation is how much dispersion around the mean, how much I a spread around the mean. In this context, how my residual is spread around the mean.

4) Mean Absolute percentage Error (MAPE)

- MAE value can range from 0 to infinity which make it difficult to interpret the result as compared to training data.
- MAPE is equivalent to MAE but provides the error in a percentage form and therefore overcomes MAE limitations.
- MAPE might exhibit some limitations if the data point (y_i) is zero since there is division involved.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{(y_i - \hat{y}_i)}{y_i}\right|$$

5) Mean percentage Error (MPE)

→ MPE is similar to MAPE but without the absolute operation. In MAPE we cannot compare positive and negative errors as ~~sometimes~~ absolute is used.

→ MPE is useful to provide how many positive errors as compared to negative ones.

$$MPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)}{y_i}$$

6) R SQUARE (R^2)

→ R square or coefficient of determination represent the proportion of variance (of y) that has been explained by independent variables in the model.

→ If $R^2 = 80$, this means 80% of the increase in ice cream cart revenue is due to increase in temperature.